ARTIKEL

IMPLEMENTASI *DATA MINING* UNTUK PENENTUAN PERBAIKAN JALAN MENGGUNAKAN ALGORITMA C4.5 (Studi Kasus Dinas Pekerjaan Umum Bina Marga Kabupaten Nganjuk)

Oleh:

Nijamul Leily R.

13.1.03.02.0088

Dibimbing oleh :

1. **M. Rizal Arief, ST., M.Kom**
2. **Patmi Kasih, M.Kom**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

2017

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2017

Yang bertanda tangan di bawah ini:

Nama Lengkap	: Nijamul Leily R.
NPM	: 13.1.03.02.0088
Telepun/HP	: 089644141894
Alamat Surel (Email)	: nijamul.leily@gmail.com
Judul Artikel	: *IMPLEMENTASI DATA MINING UNTUK PENENTUAN PERBAIKAN JALAN MENGGUNAKAN ALGORITMA C4.5* (Studi Kasus Dinas Pekerjaan Umum Bina Marga Kabupaten Nganjuk)
Fakultas – Program Studi	: Fakultas Teknik – Teknik Informatika
Nama Perguruan Tinggi	: Universitas Nusantara PGRI Kediri
Alamat Perguruan Tinggi	: Jl. KH. Achmad Dahlan 76 Kediri

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;

b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 8 Agustus 2017
Pembimbing I	Pembimbing II	Penulis,
M. Rizal Arief, ST., M.Kom 0716027504	Patmi Kasih, M.Kom. 0701107802	Nijamul Leily R. 13.1.03.02.0088

IMPLEMENTASI *DATA MINING* UNTUK PENENTUAN PERBAIKAN JALAN MENGGUNAKAN ALGORITMA C4.5 (Studi Kasus Dinas Pekerjaan Umum Bina Marga Kabupaten Nganjuk)

Nijamul Leily R.
13.1.03.02.0088
FT – Teknik Informatika
nijamul.leily@gmail.com
M. Rizal Arief, ST., M.Kom dan Patmi Kasih, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Jalan merupakan prasarana infrastruktur dasar yang dibutuhkan manusia untuk dapat melakukan pergerakan dari suatu lokasi ke lokasi lainnya dalam rangka pemenuhan kebutuhan. Permasalahan yang muncul apabila banyaknya jalan yang rusak, diakibatkan semakin banyaknya kendaraan melewati suatu ruas jalan yang usianya sudah melewati batas rencana. Selain itu permasalahan penanganan jalan dalam menentukan tindakan perbaikan jalan yang kurang tepat. Untuk mengatasi masalah tersebut, maka dibuatlah suatu sistem yang dapat digunakan untuk membantu dalam proses penentuan perbaikan jalan. Terdapat empat kriteria dasar yang digunakan yaitu panjang ruas jalan, status fungsi, tipe permukaan serta kondisi kerusakan jalan.

Algoritma C4.5 merupakan algoritma yang digunakan untuk menghasilkan pohon keputusan yang mampu mengklasifikasi suatu obyek. Dalam penelitian ini, dengan menggunakan metode C4.5 didapatkan suatu aturan (rule) penentuan perbaikan jalan berdasarkan kriteria kerusakan jalan. Metode ini dimulai dari mencari nilai gini untuk mengukur jumlah informasi yang ada pada atribut. Kemudian melakukan perhitungan terhadap gain untuk mengukur seberapa baik suatu atribut memisahkan training example ke dalam kelas target. Selanjutnya menghitung split information dan gain ratio. Atribut dengan informasi tertinggi dipilih sebagai root, sehingga didapat suatu pohon keputusan.

Hasil penelitian menunjukkan bahwa sistem yang dibangun berhasil menerapkan algoritma C4.5 dalam proses perhitungan dengan mendapatkan rule yang digunakan untuk menentukan penanganan yang tepat pada jalan yaitu pemeliharaan atau rekonstruksi. Berdasarkan simpulan hasil penelitian ini, disarankan pengembangan lebih lanjut yang dapat menyempurnakan perhitungan algoritma agar menampilkan hasil klasifikasi dengan atribut yang lebih banyak.

Kata Kunci : klasifikasi, data mining, C4.5, perbaikan jalan.

I. LATAR BELAKANG

Jalan merupakan prasarana infrastruktur dasar yang dibutuhkan manusia untuk dapat melakukan pergerakan dari suatu lokasi ke lokasi lainnya dalam rangka pemenuhan kebutuhan. Keberadaan infrastruktur jalan yang memadai sangat diperlukan. Selain untuk kemajuan perekonomian, pembangunan jalan sangat diperlukan untuk keselamatan dan kenyamanan pengendara. Apabila ada jalan yang rusak hal itu berbahaya dan bisa mengakibatkan kecelakaan.

Permasalahan yang muncul adalah apabila banyaknya jalan yang rusak, diakibatkan oleh semakin banyaknya kendaraan yang melewati suatu ruas jalan yang usianya sudah melewati batas rencana. Selain itu permasalahan penanganan jalan dalam menentukan tindakan perbaikan jalan yang kurang tepat.

Berdasarkan Pasal 13 UU No. 38 Tahun 2004,Pemerintah dan Pemerintah Daerah sebagai penyelenggara jalanmempunyai kewajiban wajib memprioritaskan pemeliharaan, perawatan dan pemeriksaan jalan secara berkala untuk mempertahankan tingkat pelayanan jalan sesuai dengan standar pelayanan minimal yang ditetapkan. Departemen Pekerjaan Umum, dalam hal ini Direktorat Jenderal Bina Marga merupakan organisasi pemerintah yang bertanggung jawab dalam penyelenggaraan infrastruktur jalan di Indonesia, salah satunya adalah Dinas Pekerjaan Umum Bina Marga Kabupaten Nganjuk.

Dinas Pekerjaan Umum Bina Marga Kabupaten Nganjuk memiliki tanggungjawab dalam melakukan perencanaan dan pemeliharaan infrastruktur di wilayah Kabupaten Nganjuk. Untuk membantu Dinas Bina Marga Kabupaten Nganjuk dalam menentukan perbaikan jalan dapat dilakukan dengan analisa data jalan, dan selanjutnya mengolah data tersebut untuk dapat menentukan jalan tersebut diperbaiki dengan cara pemeliharaan atau rekonstruksi. Analisa dan pertimbangan tersebut dapat dilakukan dengan menggunakan algoritma C4.5. Dimana algoritma ini digunakan untuk menentukan aturan-aturan berdasarkan kriteria penyebab kerusakan jalan. Kriteria yang digunakan dalam algoritma ini adalah panjang ruas jalan, status fungsi jalan yaitu JKP 4 (Jalan Kolektor Primer), JLP (Jalan Lokal Primer), J.Ling_S (Jalan Lingkungan Sekunder), J.Ling_P (Jalan Lingkungan Primer), akses ke jalan nasional, propinsi dan kota, kemudian tipe permukaan jalan berupa hotmix, aspal,

batu, serta kondisi kerusakan jalan yaitu ringan, sedang, dan berat.

Karena banyaknya jalan yang rusak, maka pemerintah harus dapat

menentukan tindakan yang tepat dalam penanganan perbaikan jalan. Dengan perbaikan jalan seperti pemeliharaan atau rekonstruksi. Untuk mengatasi masalah tersebut, maka diperlukan suatu sistem yang dapat digunakan untuk membantu dalam proses penentuan perbaikan jalan, sehingga mampu mempermudah pemerintah dalam melakukan pemeliharaan jalan.

II. METODE

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Dalam algoritma C4.5 untuk membangun pohon keputusan hal pertama yang dilakukan yaitu memilih atribut sebagai akar, kemudian dibuat cabang untuk tiap-tiap nilai di dalam akar tersebut. Langkah berikutnya yaitu membagi kasus dalam cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama (Utari, 2015).

Penghitungan nilai gini untuk mendapatkan nilai gain, terlebih dahulu kita harus mengetahui parameter lain yang mempengaruhi nilai gain, dimana parameter ini sangat diperlukan untuk mendapatkan nilai gain. Parameter tersebut adalah gini. Parameter ini sering digunakan untuk mengukur heterogenitas suatu kumpulan sampel data. Secara matematis nilai gini dapat dihitung dengan menggunakan formula sebagai berikut :

$$Gini\ (S) = 1 - \sum_{i=1}^{n} \left(\ \mathrm{p}(i|s)^2 \ \right)$$

Keterangan :

n : jumlah nilai yang ada pada atribut target (jumlah kelas)

$p(i|s)$: frekuensi relatif class i pada node s

Jika kita sudah mendapatkan nilai gini, maka langkah selanjutnya adalah melakukan perhitungan terhadap gain. Berdasarkan perhitungan matematis gain dari suatu atribut A dapat diformulasikan sebagai berikut:

$$\mathrm{Gain}(S, A) = Gini \ \sum_{i=1}^{n} \left(\ \frac{Sv}{S} * Gini(Sv) \right)$$

Keterangan :

A : atribut

V : menyatakan suatu nilai yang mungkin untuk atribut A

Values (A) : himpunan nilai-nilai yang mungkin untuk atribut A

$|Sv|$: jumlah sampel untuk nilai V

|S| : jumlah seluruh sampel data

Gini (S): gini untuk sampel-sampel yang memiliki nilai V

Untuk menghitung gain ratio kita perlu ketahui suatu term baru yang disebut

split information. Split information dihitung dengan formula sebagai berikut:

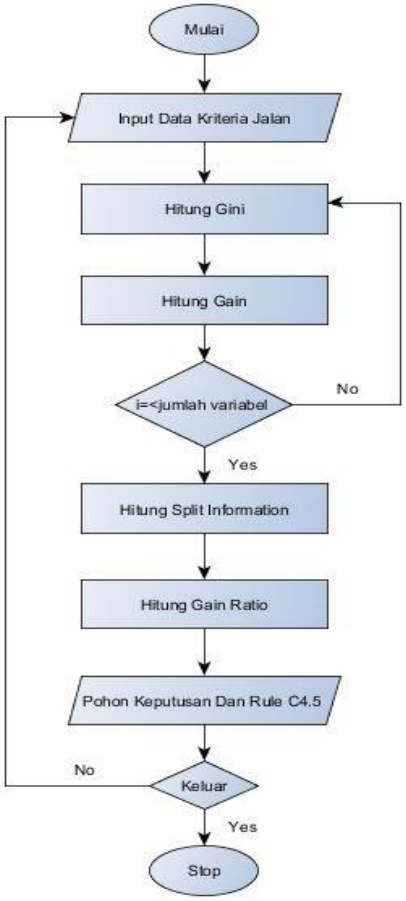
$$Split\ information = -\sum_{i=1}^{n}\left(\frac{|Sv|}{|S|} * log_2 \frac{|Sv|}{|S|} \right)$$

Selanjutnya gain ratio dihitung dengan cara :

$$Gain\ ratio = \frac{Gain}{Split\ information}$$

Proses penghitungan selesai dan mendapatkan hasil berupa pohon keputusan. Untuk melakukan proses pengolahan data perbaikan jalan maka perlu digambarkan tahapan kerja yang dapat dilakukan pada algoritma C4.5.

Berikut adalah *flowchart* dari algoritma C4.5 :



Flowchart Penghitungan Algoritma C4.5

Pada perhitungan algoritma C4.5 dibutuhkan *dataset* yang tidak berupa angka, namun jika data berupa angka harus diubah dahulu dengan ketentuan khusus. Berikut adalah tabel *dataset riil* kerusakan jalan:

Panjang (m)	Kondisi Kerusakan	Akses ke Jalan	Tipe Permukaan	Status Fungsi	Goal
1.791	ringan	nasional	hotmix	JKP 4	pemeliharaan
16.770	berat	nasional	aspal	JKP 4	rekonstruksi
8.576	ringan	kota	hotmix	JKP 4	pemeliharaan
2.500	sedang	propinsi	aspal	JLP	rekonstruksi
750	berat	kota	aspal	J.Ling_S	rekonstruksi
250	ringan	kota	hotmix	J.Ling_P	pemeliharaan
450	ringan	kota	hotmix	J.Ling_P	pemeliharaan
4.930	sedang	kota	hotmix	JKP 4	pemeliharaan
1.137	sedang	kota	aspal	J.Ling_S	pemeliharaan
1.240	sedang	kota	aspal	J.Ling_S	pemeliharaan

Dataset yang dihasilkan kemudian disesuaikan sebagai berikut:

Panjang (m)	Kondisi Kerusakan	Akses ke Jalan	Tipe Permukaan	Status Fungsi	Goal
pendek	ringan	nasional	hotmix	JKP 4	pemeliharaan
panjang	berat	nasional	aspal	JKP 4	rekonstruksi
panjang	ringan	kota	hotmix	JKP 4	pemeliharaan
pendek	sedang	propinsi	aspal	JLP	rekonstruksi
pendek	berat	kota	aspal	J.Ling_S	rekonstruksi
pendek	ringan	kota	hotmix	J.Ling_P	pemeliharaan
pendek	ringan	kota	hotmix	J.Ling_P	pemeliharaan
panjang	sedang	kota	hotmix	JKP 4	pemeliharaan
pendek	sedang	kota	aspal	J.Ling_S	pemeliharaan
pendek	sedang	kota	aspal	J.Ling_S	pemeliharaan

Keterangan:

Jalan dikatakan pendek apabila kurang dari/sama dengan 2600m dan dikatakan panjang apabila lebih dari 2600m.

Berikut adalah perhitungan dengan menggunakan algoritma C4.5, dimulai dengan menghitung jumlah masing-masing kriteria kerusakan jalan.

NODE	Kriteria	Subkriteria	TOTAL	PEM	REK
1	Jumlah		365	294	71
panjang jalan					
		pendek	196	186	10
		panjang	169	108	61
kondisi kerusakan					
		ringan	223	223	0
		sedang	114	71	43
		berat	28	0	28
akses ke					
		kota	282	230	52
		propinsi	42	34	8
		nasional	41	30	11
tipe permukaan					
		aspal	133	73	60
		hotmix	232	221	11
status fungsi					
		JKP 4	72	64	8
		JLP	57	28	29
		J.Ling_P	155	132	23
		J.Ling_S	81	70	11

Selanjutnya menghitung nilai *gini*, *gain*, *split information* dan *gain ratio* dari tiap kriteria diatas.

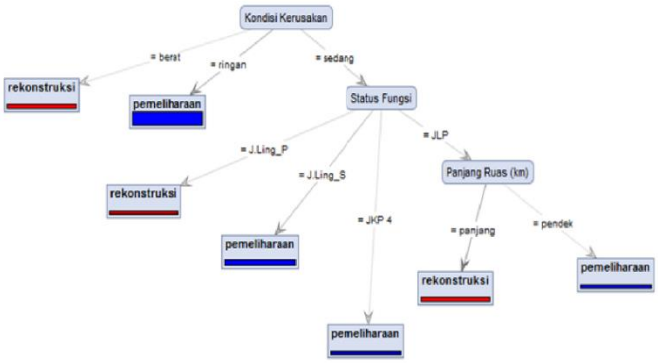
Kriteria	Subkriteria	GINI	GAIN	SPLIT INF	GAIN RATIO
Jumlah		0.389			
panjang jalan			0.123	0.996	0.124
	pendek	0.097			
	panjang	0.461			
kondisi kerusakan			0.242	1.243	0.195
	ringan	0.000			
	sedang	0.470			
	berat	0.000			
akses ke			0.077	1.001	0.077
	kota	0.301			
	propinsi	0.308			
	nasional	0.393			
tipe permukaan			0.151	0.946	0.160
	aspal	0.495			
	hotmix	0.090			
status fungsi			0.113	1.887	0.060
	JKP 4	0.198			
	JLP	0.500			
	J.Ling_P	0.253			
	J.Ling_S	0.235			

Dari perhitungan diatas didapat nilai *gain ratio* terbesar yaitu kriteria kondisi kerusakan, yang akan digunakan sebagai *root*, dan penghitungan akan diulang lagi sampai menghasilkan pohon keputusan.

III. HASIL DAN KESIMPULAN

A. HASIL

Setelah semua penghitungan selesai, didapat pohon keputusan dan *rule* sebagai berikut :



Dari pohon keputusan diatas dapat disimpulkan *rule* sebagai berikut:

1. *IF* Kondisi Kerusakan = Berat *THEN* Rekonstruksi
2. *IF* Kondisi Kerusakan = Ringan *THEN* Pemeliharaan
3. *IF* Kondisi Kerusakan = Sedang *AND* Status Fungsi = J.Ling_P *THEN* Rekonstruksi
4. *IF* Kondisi Kerusakan = Sedang *AND* Status Fungsi = J.Ling_S *THEN* Pemeliharaan
5. *IF* Kondisi Kerusakan = Sedang *AND* Status Fungsi = JKP *THEN* Pemeliharaan
6. *IF* Kondisi Kerusakan = Sedang *AND* Status Fungsi = JLP *AND* Panjang Ruas = Panjang *THEN* Rekonstruksi
7. *IF* Kondisi Kerusakan = Sedang *AND* Status Fungsi = JLP *AND* Panjang Ruas = Pendek *THEN* Pemeliharaan

B. KESIMPULAN

1. Hasil perhitungan dengan Algoritma C4.5 berupa *rule* dapat digunakan dalam menentukan penanganan yang tepat pada kerusakan jalan.
2. Sistem yang dapat membantu petugas Bina Marga dalam menentukan perbaikan jalan dibangun dengan menerapkan konsep kerja *Data Mining* dan Algoritma C4.5 sebagai metode pendukungnya.

IV. DAFTAR PUSTAKA

[1] Abidin, A. Z. 2011. *Implementasi Algoritma C4.5 untuk Menentukan Tingkat Bahaya Tsunami*. Seminar Nasional Informatika 2011 (semnasIF 2011). Yogyakarta: UPN "Veteran" Yogyakarta. (Online), tersedia: http://repository.upnyk.ac.id/620/1/A-4.pdf, diunduh 18 Oktober 2016 pukul 20.35 WIB.

[2] Arista, N. 2016. Penerapan *Algoritma ID3 dan C4.5 dalam Menemukan Hubungan Data Awal Masuk Mahasiswa dengan Prestasi Akademik.* (Online), tersedia: http://jurnal.umrah.ac.id.pdf, diunduh 18 Oktober 2016 pukul 20.22 WIB.

[3] Ferdiansyah, N. 2015. *Sistem Pendukung Keputusan Penentuan Prioritas Perbaikan Jalan Di Dinas Bina Marga Kabupaten Cirebon Dengan Metode Topsis.* Seminar Nasional Teknologi Informasi dan Multimedia 2015. Yogyakarta: STMIK AMIKOM Yogyakarta. (Online), tersedia: http://ojs.amikom.ac.id/index.php, diunduh 19 Oktober 2016 pukul 19.47 WIB.

[4] Han, J. Kember, M. 2006. *Data Mining: Concepts and Techniques, 2nd edition. The Morgan Kaufman series in Data Management System*. Jim Grey, series Editor. (Online), tersedia : http://web.engr.illinois.edu/~hanj/bk2/slidesindex.htm, diunduh 16 Oktober 2016 pukul 20.21 WIB.

[5] Kadir, A. 2009. *Membuat Aplikasi Web dengan PHP+Database MySQL*. Yogyakarta: Andi Offset. (Online), tersedia: https://scholar.googlecom/citations?user=AusNQ7MAAAJ&hl=ja, diunduh 18 Oktober 2016 pukul 20.46 WIB.

[6] Kusrini. 2006. *Sistem Pakar: Teori dan Aplikasi*. Yogyakarta: Andi Offset. (Online), tersedia: https://books.google.co.id/books/about/algoritma_data_mining.html?hl=id&id=-Ojclag73O8C, diunduh 16 Oktober 2016 pukul 19.55 WIB.

[7] Larose D, T., 2005, *Discovering knowledge in Data: an Introduction to Data mining*. Jhon Wiley & Sons Inc. (Online), tersedia: http://onlinelibrary.wiley.com/book/10.1002/0471687545, diunduh 16 Oktober 2016 pukul 21.30 WIB.

[8] Santosa, Budi. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu.Yogyakarta. diunduh 7 Juni 2017 pukul 13.45 WIB.

[9] Tan, dkk. 2006. *Intoduction to Data mining*. Person Education, Inc. (Online), tersedia: http://www-user.cs.umn.edu, diunduh 16 Oktober 2016 pukul 21.22 WIB.

[10] Turban, E. 2005. *Decision Support System and Intelligent System* 7th *Ed*. New Jersey: Pearson Education. (Online), tersedia: https://www.amazon.com/Decision-Support-System-Intelligent-6th/dp/0130894656, diunduh 16 Oktober 2016 pukul 20.42 WIB.

[11] Utari, S. 2015. *Implementasi Metode C4.5 untuk Menentukan Guru Terbaik pada SMK 1 Percut Sei Tuan Medan, Volume. IX No.3*. Medan. (Online), tersedia: http://www.e-jurnal.com/2016/09/implementasi-metode-c45-untuk.hyml?m=1, diunduh 19 Oktober 2016 pukul 19.33 WIB.

[12] Uzzahroh, A. (2014). Sistem Pendukung Keputusan Penentuan Lokasi Perbaikan Jalan dengan Metode Analytical Hierarchy Proses (AHP). Yogyakarta: UIN Sunan Kalijaga. (Online), tersedia: http://digilib.uin-suka.ac.id/13647/, diunduh 22 Oktober 2016 pukul 21.52 WIB.